DOCUMENT RESUME

ED 185 518 CS 005 382

AUTHOR	Gil, Doron; And Others
TITLE	The Impact of Training on Diagnosis Consistency. Research Series No. 67.
INSTITUTION	Michigan State Univ., East Lansing. Inst. for Research on Teaching.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Jan 80
CONTRACT	400-76-0073
NOTE	63p.; Best copy available.
AVAILABLE FROM	Institute for Research on Teaching, College of Education, Michigan State University, 252 Erickson Hall, East Lansing, MI 48824 ($5.00)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Diagnostic Tests; Graduate Study; Higher Education; Pretests Posttests; Reading Consultants; *Reading Diagnosis; *Reading Difficulties; *Reading Research; *Reading Teachers; Reading Tests; *Reliability; Remedial Reading; *Teacher Education

ABSTRACT

A study was conducted to investigate the degree to which subjects could be trained to improve the reliability of their diagnostic decisions about cases of reading difficulties. Twenty-eight graduate students enrolled in a reading diagnosis course were trained from G. B. Sherman's Model of Reading and Learning to Read, and practiced on simulated cases of reading difficulty. Pretests and posttests were administered, in which the students were asked to diagnose cases, write their diagnoses, and transfer their diagnoses to checklists. Statistics were run to calculate the students' diagnostic agreement with each other. The students' diagnostic agreement approximately doubled from pretests to posttests, indicating that subjects can be trained to improve the reliability of their diagnostic decisions about cases of reading difficulties. (Eleven appendixes include copies of materials used in the course.) (Author/MKM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED185518

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Research Series No. 67

THE IMPACT OF TRAINING ON DIAGNOSTIC CONSISTENCY

Doron Gil, Ruth M. Polin,
John F. Vinsonhaler and Joel VanRoekel

BEST COPY AVAILABLE

Published By

The Institute for Research on Teaching
252 Erickson Hall
Michigan State University
East Lansing, Michigan 48824

January 1980

CS005382

This work is sponsored in part by the Institute for Research on Teaching, College of Education, Michigan State University. The Institute for Research on Teaching is funded primarily by the Program for Teaching and Instruction of the National Institute of Education, United States Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the National Institute of Education. (Contract No. 400-76-0073)

Institute for Research on Teaching

Teachers' thoughts and decisions are the focus of studies currently under way at Michigan State University's Institute for Research on Teaching (IRT). The IRT was founded in April 1976 with a $3.6 million grant from the National Institute of Education. That grant has since been renewed, extending IRT's work through September 1981. Funding is also received from other agencies and foundations. The Institute has major projects investigating teacher decision-making, including studies of reading diagnosis and remediation, classroom management strategies, instruction in the areas of language arts, reading, and mathematics, teacher education, teacher planning, effects of external pressures on teachers' decisions, socio-cultural factors, and teachers' perceptions of student affect. Researchers from many different disciplines cooperate in IRT research. In addition, public school teachers work at IRT as half-time collaborators in research, helping to design and plan studies, collect data, and analyze results. The Institute publishes research reports, conference proceedings, occasional papers, and a free quarterly newsletter for practitioners. For more information or to be placed on the IRT mailing list please write to : The IRT Editor, 252 Erickson, MSU, East Lansing, Michigan 48824.

Director: Judith E. Lanier

Associate Directors: Lawrence W. Lezotte and Andrew C. Porter

Editorial Staff:

Lawrence W. Lezotte, coordinator of Communications/Dissemination
Linda Shalaway, IRT editor
Janet Flegg, assistant editor

(Lee S. Shulman, co-director with Judith E. Lanier from 1976-1978, and director in 1979, is on a one-year leave at the Center for Advanced Study in the Behavioral Sciences, Stanford, California.)

Acknowledgements

The authors wish to acknowledge the contributions of the following to the conduct of this study: George Sherman, whose cooperation and expertise enabled us to follow the progress of his students in Education 830E, and Christian Wagner, who, with George Sherman, developed materials for the course.

THE IMPACT OF TRAINING ON DIAGNOSTIC CONSISTENCY

Doron Gil, Ruth M. Polin, John F. Vinsonhaler,
and Joel VanRoekel[1]

Whether or not clinicians reach similar or different diagnoses when interacting with a given case is an important issue for all professions in which clinicians must make decisions about their clients (Vinsonhaler, Note 1). The importance of this issue stems from the assumption that diagnosis and remediation are interrelated. For remediation to be effective, it should direct itself to the client's problems. This will be achieved only if the diagnosis is reliable.

The question of diagnostic agreement is especially important in medicine, where incorrect diagnosis, yielding improper patient management, can result in permanent loss to the patient, and, in some cases, even in death.

Because of the issue's seriousness, many studies have been conducted to investigate inter-physician agreement. Several investigations have revealed that "a surprising and clinically important degree of inaccuracy is to be expected in the interpretation or evaluation of many clinical and laboratory procedures used in everyday practice" (Garland, 1959). For

---

[1]Doron Gil is a research associate with IRT's Clinical Studies Project. Ruth M. Polin is the group's data processing coordinator, John F. Vinsonhaler is the project coordinator, and Joel VanRoekel, a former Clinical Studies intern, is a reading specialist with the Waverly School District in Lansing.

example, writes Garland,

> experienced interpreters of a series of plain chest roentgenograms[2] are apt to miss about 30% of those films positive for roentgenologic evidence of disease, and to over-read about 2% of those negative for disease. Further, in evaluating pairs of serial roentgenograms for alterations in the status of disease, one experienced physician is apt to disagree with another in about one third of the cases, and (on review) to disagree with himself on one fifth of them. Comparable degrees of error occur in many forms of clinical practice. Indeed, if all branches of medicine could be tested, the phenomenon would probably be found quite universal.

Studies by other researchers support Garland's statements. In an examination of the reliability of chest radiography in the diagnosis of pulmonary lesions, Yerushalmy (1955) found that an individual physician agreed with other radiologists on only about 30% of the X rays. Yerushalmy concluded that "observer error in chest radiography is of such magnitude as to create special problems in diagnosis and treatment." In another study, Paton (1957) found that "the accuracy rate in the diagnosis of myocardial infarction is surprisingly low, only 44%."

Given the importance placed on diagnostic agreement in medicine, it seems logical that this would also be an important consideration in diagnosis of reading difficulties in children.

Clinicians' agreement in reading diagnosis and remediation has been the focus of intensive investigation conducted by IRT's Clinical Studies Project. The theoretical base for this investigation involves a theory of clinical problem-solving behavior, initially originated in medical research as the Inquiry Theory by Elstein and Shulman (Elstein, Shulman, & Sprafka, 1978), and reformulated in education by Vinsonhaler and Wagner (Vinsonhaler, Wagner, & Elstein, 1977).

---

[2] A roentgenogram is a photograph made with X rays.

The Inquiry Theory postulates that clinical problem-solving behavior is a function of clinicians' memories and strategies, which form clinicians' decision-making processes as they diagnose cases. The memories and strategies are determined, in part, by the training clinicians receive. The more similar their training, the more similar their memories and strategies will be, and the more closely they will agree with each other when diagnosing the same cases. To the extent that their training varies, clinicians will employ different "memories" or "models" of the reading process, and will show limited agreement in their diagnostic judgments.

The Agreement Corollary, derived from the Inquiry Theory, states that the more similar clinicians' memories and strategies are, the higher their diagnostic agreement will be.

To investigate the Inquiry Theory and the Agreement Corollary, several observational studies were conducted by the Clinical Studies group. The subjects were reading clinicians, learning disability resource teachers, and classroom teachers. They were asked to interact with simulated reading and learning difficulty cases and to arrive at diagnostic judgments about the cases' problems. (For a more detailed description of these studies, see Gil, Hoffmeyer, VanRoekel, & Weinshank (Note 2).)

Part of the data analysis for these studies focused on the reliability (or consistency) of clinicians' diagnoses in comparison with themselves and other clinicians. A Phi Coefficient correlation was applied to calculate this reliability. (For a discussion of the Phi Coefficient, see Appendix J.) Table 1 summarizes results for five of these studies and a similar study done with physicians (Norman, Barrows, Neufeld, & Feightner, Note 3).

As Table 1 illustrates, results of these five studies indicate that little agreement existed among the reading clinicians' diagnoses.

Table 1

Mean Phi Correlations of Observational Studies 1977-1979

Study	0877	0878.1	0878.2	0878.3	0879.2	General/Internal Medicine
Subject	Reading Clinicians	Learning Disability & Reading Clinicians	Classroom Teachers	Reading Clinicians	Reading Clinicians	Medical Doctors
No. of Subjects	8	20	10	8	8	37
No. of Observations	24	20	20	24	24	58
Mean Phi Intercorrelations	-.07	-.05	-.04	.11	.16	.34
Standard Deviation	.17	.11	.13	.09	.10	.28

Possible explanations for these results could be that these low correlations resulted from the varied training backgrounds clinicians receive in diagnosis, and/or that they lack or possess different comprehensive models of reading.

To investigate the impact of training on clinicians' diagnoses, a training study involving 36 students enrolled in a graduate reading diagnosis course was conducted in the summer of 1977 (Sherman, Weinshank, & Brown, Note 4). In addition to the traditional presentation of course content by the instructor, the students interacted with simulated cases of reading difficulties. Pre- and posttests were used to measure the students' diagnostic performance scores, which increased significantly by the end of the course. (Analyses were run to find the degree to which they agreed with the instructor's diagnoses for the cases. No analyses were run to find the diagnostic agreement of a student with him/herself.)

The purpose of the study reported here was to further investigate the impact of training on the reliability of clinical decisions in reading

diagnosis. More specifically, these were the objectives:

1. To determine the degree to which teaching reading diagnosis from a specific yet comprehensive model (Model of Reading and Learning to Read, MORAL[3]) improves diagnostic agreement.

2. To determine the degree to which practice on simulated cases of reading difficulties with corrective feedback results in diagnostic agreement, and

3. To determine the degree to which the diagnostic agreement of clinicians improves through the use of such decision aids as the channeling of subjects' thinking, diagnostic write-ups, and consideration of four vital signs of reading (instant word recognition, decoded word recognition, fluent text segmentation, and retentive comprehension.

Methods

The subjects in this study were 28 students enrolled in the 1979 summer graduate level reading diagnosis course (ED 830E) taught by George Sherman at Michigan State University.

The course, lasting five weeks (30 hours of class), involved two major components derived from the study's objectives. First, the students received instruction based on a Model of Reading and Learning to Read (see course syllabus, Appendix A). Second, the students were given opportunities to practice four simulated cases of reading difficulties, generate diagnostic decisions about the cases, write them up, and receive the instructor's feedback on their written diagnoses. These were training cases, different from the ones the students diagnosed in pre- and posttests. Third, students were given decision aids (i.e., Diagnostic/Prescriptive summary forms) to guide their interactions with the simulated cases. The overall design of the study is summarized in Table 2.

Sherman's Model of Reading and Learning to Read postulates that clinical

[3]This model was developed by George B. Sherman, Institute for Research on Teaching, Michigan State University.

Table 2

Overall Design of the Study

Subjects	Pretest	Class Content	Posttest
Four groups of seven students each	1. Instructions 2. Interaction with a simulated case (1, 2, 3, or 4) Max = 45 min 3. Diagnostic write-up 4. a. Date Base Checklist b. Causality Checklist	1. Training on the MORAL 2. Practice on replicate case with instructor's feedback	1. Instructions 2. Interaction with same case as in pretest (1,2,3, or 4) Max = 45 min 3. Diagnostic write-up 4. a. Data Base Checklist b. Causality Checklist c. Diagnostic Checklist

decisions with regard to reading diagnosis and remediation will be more consistent and effective if the clinician considers students' performance levels on four vital signs of reading (Sherman, Note 5):

1. *Instant word recognition*, which is the student's ability to recognize words without hesitation as measured by relative size of sight word vocabulary;

2. *Decoded word recognition*, which is the student's ability to recognize unfamiliar words through the use of graphemic redundancies;

3. *Fluent text segmentation*, which is the student's ability to translate groups of words into cognitive structures with semantic meaning and to briefly retain these structures; and

4. *Retentive comprehension*, which is the student's ability to retain the results of perceptive comprehension and to organize these results into more complex semantic structures suitable for long term retention, complex inference, and problem solving.

See Appendix B for a chart detailing the causal factors which affect each of these vital signs.

Data Collection

On the first day of class, students were randomly assigned to one of four groups (seven students in each group, N = 28). The four groups were to be identical on all variables (e.g., training on the MORAL and practice on simulated cases) but different with regard to the simulated cases they were asked to diagnose for pre- and posttests. (These four simulated cases, although identical in format, represented different reading problems.)

The students were given complete directions for diagnosing the simulated cases on their pre- and posttests.

First, they were given a simulated case with which to interact. Each simulated case included various forms of information about a child's reading performance, such as the child's background, audio recordings of the child reading aloud, test records, test descriptions, and test results. The students listened to recorded instructions about their task, and also received printed instructions (see Appendix C).

After receiving the instructions, the students observed the initial contact information about the case, which included a short summary about the child's reading performance (e.g., "The child is eight years old and reading at fourth-grade level."). The students then were given 45 minutes to collect as many cues (items of information) about the case as they wished. They were asked to list all cues selected on a record form in the order in which they were collected (see Appendix D).

At the end of 45 minutes, the students were asked to write their diagnoses about the case according to a diagnostic/prescriptive summary form developed by George Sherman (see Appendix E). This form channelled their thinking and diagnostic write-up along the four vital signs of

reading and their causal factors.

After completing the diagnostic write-up, students were asked to match their written diagnoses with diagnostic categories listed on two different checklists. This standardized the students' vocabulary for comparison and data analysis.

The rationale for requiring students to complete these checklists at the end of their diagnostic session, after going through the process of diagnostic write-up and transfer, is that students may have operated from a response set,[2] marking all items in the checklists whether these represented the case or not. Second, asking students to respond only to a checklist channels their thinking along certain problem areas and limits their judgment to decide, independently, what the child's reading problems are. These final decisions, after all, were the focus of this study.

The first checklist (termed the MORAL Data Base Checklist) listed 49 statements about a child's reading status (see Appendix F). The students were instructed to mark only those categories which they had mentioned in their diagnostic write-up. For example, if a student mentioned in his/her diagnostic write-up "the child lacks visual memory," s/he was supposed to check category 7 (visual memory of word forms) as inadequate on the checklist.

The same procedures applied for the second checklist, termed the MORAL Causality Checklist, (see Appendix G). It listed 25 statements indicating various causes of a child's poor reading for the four vital signs. If, for example, a student mentioned in his/her diagnostic write-up that the child has poor instant word recognition because of lack of practice, s/he should have checked the first statement, "Inadequate instant word recognition is partially caused by insufficient independent reading practice,"

[2] A response set is the psychological tendency to respond to all items on a checklist.

on the causality checklist.

The students were allowed a few minutes to work with each checklist. All diagnoses, record forms for the cues, and checklists for each student were filed and assigned a random number (1-28) to facilitate data analysis.

Identical procedures were followed in the posttest session, which took place on the last day of class. The students received the same case they diagnosed for their pretest. This time, however, the students were asked to work on a third, additional checklist (see Appendix H). This checklist (termed a Diagnostic Checklist) required the students to transfer their diagnostic write-ups to a list of categories pertaining to the reading act (these included Physical, Affective, Family, Perception, Language, School, Structural Analysis, Oral Reading, Silent Reading, Phonetic Analysis, Word Recognition, and Comprehension). For each category the students were asked to indicate whether the child's reading behavior was a strength (S), which contributed to the reading act; a weakness (W), which inhibited the reading act; or whether they were making a mere observation (O) (i.e., no value judgment was made). If, for example, the student mentioned in his/her diagnostic write-up that the child's rate of reading was poor, s/he should have marked "Oral Reading: Rate" as "W" (weakness). As with the two previous checklists, the students were instructed to check only those categories which corresponded to those appearing in their own diagnostic write-ups.

The students' responses to this checklist allowed us to study the transfer of training on the Model of Reading and Learning to Read to typical diagnoses. Such a transfer is important in order for the students to actually apply their diagnostic skills to reading problems encountered in actual classrooms.

Reading Diagnosis Training

The formal classroom instruction component for this study was conducted in three-hour blocks, twice weekly, for five weeks. The course topics and their order of presentation were governed by Sherman's Model of Reading and Learning to Read. A handout containing the model in matrix form (see Appendix B) was distributed to students on the first day of class to assist them in placing the various aspects of the course within the perspective of the model. To assist students in grasping the important aspects of a vital sign or causal factor, various demonstrations were planned. These demonstrations occurred both during and/or following a lecture. They included the administration, scoring and interpretation of various measures used to assess the vital signs of reading, the effect of performances on specific vital signs as they related to the total reading act, alternative methods of assessment, and more. A simulated case similar to those used for the pretest served as the basis for many of the demonstrations.

In addition to the six hours of classroom instruction, students were required to interact with one practice simulated case each week (four in all). These interactions were designed to promote student application of the information covered during lectures and demonstrations. The practice simulated cases were available for use throughout the day and could be used by an individual or a small group of students simultaneously. All practice simulated cases were computer based; manually based simulated cases were used for pre- and posttesting. All the information in the computer version was in the same format as that in the manual version. After interacting with a practice simulated case, students were required to complete Data Base and Causality Checklists for that case.

Each practice simulated case was discussed during formal class time.

The discussion included Sherman's reaction to and interpretation of the information in the case as it related to the Model of Reading and Learning to Read. Little if any formal class time was devoted to a discussion of the two checklists because of time constraints. Upon completion of all four practice cases and nine lecture/demonstration periods, a final review of the Model of Reading and Learning was presented prior to posttesting.

Results

Data analysis focused primarily on the degree of diagnostic agreement of the students with themselves and each other. For that purpose, Phi and Porter Correlations were calculated to find the diagnostic agreement of each student on pre- and posttests on the Data Base and Causality Checklists. The mean diagnostic agreement of all students diagnosing the same case was also calculated.

These comparisons were designed to indicate the degree to which the students' written diagnoses for a case were reliable. These comparisons were made on the basis of the diagnostic checklists, rather than on the students' written diagnoses, because of problems with standardized vocabulary. For that reason, it was essential to make sure that the students had, indeed, marked on the checklists only those diagnostic categories they used in their diagnostic write-ups. Joel VanRoekel, who assisted Sherman in the course, reviewed each student's checklists to verify this. All categories that were not included in the students' initial diagnostic write-ups were discarded. For reliability purposes, this verification procedure was repeated once more on a random sample of diagnostic checklists by another member of the Clinical Studies group. The agreement score with VanRoekel's verification was .85. Data analysis was then run on all diagnostic categories which had been thus verified as being part of the students' diagnoses.

It appears that students' transfer of their written diagnoses to the checklists was affected by a response set. Most of the students marked almost all categories on the checklists, despite instructions to mark only those items which were in their own written diagnoses. Although the reasons for this phenomenon are not yet clear, we have a few hunches.

First, although the students were told that their performance on the checklists would not affect their course grade, perhaps some of them, thinking that poor performance might affect their grade anyway, marked as many categories as they could. On the other hand, it may well be that a few students, knowing that their performance on the checklists would not affect their grade, hadn't approached their task as thoughtfully as they should have. Yet another explanation could be that, at least in some cases, students didn't quite understand the checklist vocabulary, so they guessed at a few of the categories according to what they perceived was meant. Fourth, going through the checklists, students may have encountered diagnostic categories that, according to their judgment, have characterized the case (even though they have not included these categories in their diagnostic write-ups). Nevertheless, they checked these categories in the checklists. Lastly, it may be that some students just didn't know how to diagnose the case, so they marked, almost indiscriminately, as many categories as they could.

Calculating proportional agreement scores and running Phi Correlations on all categories checked by the students under these conditions would have meant inflating the reliability of the scores. For that reason, calculations were made on only those categories which were originally mentioned in the students' diagnostic write-ups (i.e., not eliminated from the checklists by VanRoekel).

Proportional Agreement

Proportional agreement is a measure of group agreement, based on those items in the checklists which were included in the student's own diagnostic write-ups, or on the list of cues collected by all the students for a case. It is determined by first defining a domain of statements (statements in the checklists or cues collected). Next, the percent of students who checked the same statements or collected the same cues about a case is calculated. This measure indicates which diagnostic statements were most frequently mentioned or which cues were more frequently collected by the students.

Proportional agreement for diagnosis. Table 3 presents the proportional agreement scores for the Data Base Checklist on pre- and posttests. The displayed results suggest that diagnostic agreement increases with training.

Table 4 presents the proportional agreement scores for the Causality Checklist on pre- and posttests. In general, there was higher proportional agreement on student scores for the posttest than the pretest; the students reached closer diagnostic agreement after training.

Proportional agreement for cues. Proportional agreement for cues was calculated in a manner similar to that used for the diagnostic checklists. First, a domain of all cues available for each simulated case (four domains in all) was prepared. Second, the proportion of students requesting each cue was determined. There were five kinds of cues: test scores, examiner's comments, test booklets, audio recordings, and test directions. The mean proportional agreement, pretest and posttest, for each of the five kinds of cues, and the net change after five weeks of training were calculated. They are presented in Table 5. The proportional agreement of the cues themselves (selected by at least 50% of the subjects) can be found in Appendix I.

Table 3

Diagnostic Statements from Data Base Checklist Mentioned Most Frequently (by at least 40% of the students) on pre- and posttests*

Statement	Simulated Case	Student Proportional Agreement							
		1		2		3		4	
		pre-test	post-test	pre-test	post-test	pre-test	post-test	pre-test	post-test
Instant word recognition--Inadequate		.71	1.00	.86	1.00	.00	.29	.71	1.00
Decoded word recognition level 1.0-3.9--Inadequate		.57	.71	.29	.57	.00	.00	.86	.86
Decoded word recognition level 4.0-8.9--Inadequate		.00	.14	.43	.57	.43	.71	.29	.71
Fluent text segmentation Instant word recognition level--Inadequate		.86	.43	.29	.43	.29	.57	.57	.57
Application of decoding skills for multisyllable word identification--Inadequate		.14	.57	.00	.71	.14	.71	.14	.57
Visual Memory of word forms--Inadequate		.57	.86	.29	.00	.00	.00	.57	.57
Use of punctuation signals in written material--Inadequate		.43	.57	.00	.14	.43	.71	.43	.14
Levels of independent contextual practice for normal development of instant word recognition achievement--Inadequate		.14	.57	.00	.71	.00	.71	.00	.43
Levels of independent contextual practice for normal development of decoded word recognition achievement--Inadequate		.00	.57	.14	.57	.00	.71	.00	.43
Levels of independent contextual practice for normal development of fluent text segmentation achievement--Inadequate		.00	.57	.00	.43	.00	.71	.14	.43
Message comprehension--Adequate		.71	.57	.14	.00	.29	.43	.71	.43
Visual chunking for decoded word recognition--Inadequate		.43	.29	.14	.43	.14	.29	.14	.71
Message comprehension--Inadequate		.00	.14	.57	1.00	.14	.43	.14	.43
Auditory acuity for letter and phonemic feature discrimination--Inadequate		.00	.00	.43	.86	.29	.00	.00	.00
Instant word recognition--Adequate		.00	.00	.00	.00	.57	.43	.14	.00

*Ordered by frequency in posttest

Note: 0.0. indicates that none of the students mentioned this category.

Table 4

Student Proportional Agreement on Diagnostic Statements from Causality Checklist Mentioned Most Frequently (by at least 40% of the students) on Pre- and Posttests.*

Simulated Case	1		2		3		4	
	pre-test	post-test	pre-test	post-test	pre-test	post-test	pre-test	post-test
Inadequate IWR is partially caused by insufficient independent reading practice	.14	1.00	.00	.71	.14	.71	.14	1.00
Inadequate DWR is partially caused by lack of practice	.00	.71	.00	.71	.00	.71	.14	.57
Inadequate FMS (at IWR level) is partially caused by inappropriate use of punctuation	.29	.57	.00	.14	.71	.86	.29	.14
Inadequate FMS (at grade placement) is partially caused by slow DWR	.71	.57	.14	.57	.14	.29	.14	.57
Inadequate FMS (at grade placement) is partially caused by inadequate IWR	.29	.57	.29	.57	.14	.14	.14	.14
Inadequate DWR is partially caused by lack of independent application of known decoding skills	.00	.86	.00	.14	.14	.71	.00	.43
Inadequate DWR is partially caused by inadequate Sound Symbol Association for vowel phonograms	.14	.57	.00	.29	.29	.00	.57	.57
Inadequate IWR is partially caused by inadequate visual memory	.43	.71	.14	.00	.00	.14	.57	.71
Auditory acuity deficit impairs overall phonics skills: both generalization and transfer performance	.00	.00	.57	1.00	.29	.00	.00	.00
Inadequate DWR is partially caused by inadequate blending of word parts	.57	.29	.29	.14	.29	.00	.43	.29
Inadequate IWR is partially caused by inadequate visual acuity	.14	.00	.14	.00	.57	.43	.14	.00
Inadequate IWR is partially caused by inadequate visual discrimination	.29	.29	.43	.00	.00	.00	.00	.57

*Ordered by frequency in posttest

Note: 0.0. indicates that none of the students mentioned this category.

Key

IWR = instant word recognition
DWR = decoded word recognition
FMS = fluent text segmentation

Table 5

Mean Proportional Agreement on Type of Cue Requested

	Simulated Case	Test Scores			Examiner Comments			Test Booklets			Audio Recording			Test Directions		
		Pre-test	Post-test	Change	Pre-test	Post-test	Change	Pre-test	Post-test	Change	Pre-test	Post-test	Change	Pre-test	Post-test	Change
Mean	1	.50	.27	-.23	.43	.28	-.15	.36	.54	+.18	.14	.24	+.10	.14	.01	-.13
Standard Deviation		.19	.19		.16	.26		.16	.39		.13	.22		.10	.04	
Mean	2	.42	.30	-.12	.42	.24	-.18	.29	.56	+.27	.14	.57	+.43	.07	.14	+.07
Standard Deviation		.15	.23		.21	.18		.20	.39		(based on one item)			.10	.32	
Mean	3	.40	.30	-.10	.22	.24	+.02	.16	.30	+.14	.18	.25	+.07	.05	.01	-.04
Standard Deviation		.22	.19		.37	.21		.14	.30		.18	.41		.08	.04	
Mean	4	.49	.47	-.02	.50	.27	-.23	.27	.41	+.14	.16	.10	-.06	.08	.05	-.03
Standard Deviation		.17	.20		.15	.22		.25	.33		.19	.11		.16	.11	
Mean	Grand Mean	.45	.33	-.12	.39	.26	-.13	.27	.45	+.18	.16	.29	+.13	.09	.05	-.04
Standard Deviation		.05	.09		.12	.02		.08	.12		.02	.20		.04	.06	

The results show a trend away from the use of test scores and examiner's comments toward increased use of test booklets. The training, with instruction on cue collection, seemed to influence the students to study the test booklets themselves, rather than to depend on second-hand information from test scores and examiner comments. There seemed to be a slight increase in the use of audio recordings as well.

Inter-Clinician Agreement Correlation

Inter-agreement correlation is a measure of the agreement between the students on the statements made or cues collected for a given case. This statistic is calculated as follows: (1) A domain of statements (statements on the checklists or cues collected) is defined for a case; (2) Those statements appearing in an individual student's interaction with the case are identified; and (3) A two-by-two contingency table is prepared, and a correlation is calculated. The cells in the contingency table include (a) the number of statements mentioned in both encounters, (b) statements present in the first encounter (one student) but not in the second, (c) statements present in the second encounter (second student) but not in the first, and (d) statements not present in either encounter.

Two statistical procedures, the Phi and Porter Coefficients, were used to calculate the inter-agreement scores. For a more detailed discussion of these statistics, see Appendix J.

Inter-clinician agreement for diagnoses. Tables 6 and 7 present the mean inter-clinician correlation (Phi and Porter Coefficients) for the four cases on pre- and posttests for the Data Base Checklist and the Causality Checklist, respectively.

Table 6

Inter-Clinician Correlations Based on Data Base Checklist

Simulated Case	Pretest				Posttest			
	Phi	Standard Deviation	Porter	Standard Deviation	Phi	Standard Deviation	Porter	Standard Deviation
1	.32	.13	.22	.09	.39	.13	.28	.11
2	.23	.18	.15	.12	.47	.12	.34	.10
3	.15	.17	.09	.10	.31	.16	.23	.13
4	.33	.15	.22	.09	.36	.12	.26	.10
Grand Mean	.26	.08	.17	.06	.38	.07	.28	.05

Table 7

Inter-Clinician Correlations Based on Causality Checklist

Simulated Case	Pretest				Posttest			
	Phi	Standard Deviation	Porter	Standard Deviation	Phi	Standard Deviation	Porter	Standard Deviation
1	.16	.27	.17	.16	.38	.18	.39	.11
2	.14	.25	.13	.16	.37	.18	.33	.14
3	.19	.23	.16	.12	.37	.24	.34	.18
4	.14	.26	.14	.18	.39	.20	.37	.16
Grand Mean	.16	.02	.15	.02	.38	.01	.36	.03

The data summarized in Tables 6 and 7 indicate that the students' diagnostic agreement improved from pre- to posttests on both checklists. In other words, at the end of the five-week course, the students demonstrated greater agreement among themselves on cases of reading difficulties than at the beginning of the course.

The data further show that a greater improvement was obtained on the Causality Checklist (from .16 to .38) than on the Data Base Checklist (.26 to .38). As seen, students began the course with higher Phi correlations on the Data Base Checklist, and improved less on it, than was the case for the Causality Checklist (on which they doubled their scores in five weeks). Undoubtedly, the *causes* of a reading problem require more judgment and are more difficult to identify than the problem itself. Throughout the course, as students became more qualified in juding the causes underlying reading difficulties, they showed the same sophistication on both checklists.

Table 8 shows inter-clinician correlations for the Diagnosis Checklist. They are comparable to those obtained for the Data Base and Causality Checklists.

Table 8

Inter-Clinician Correlations Based on 79DX Checklist (Posttest Only)

Simulated Case	Phi Coefficient	Standard Deviation	Porter Coefficient	Standard Deviation
1	.31	.14	.21	.10
2	.29	.11	.18	.07
3	.28	.14	.19	.09
4	.33	.12	.22	.09
Grand Mean	.30	.02	.20	.02

Inter-clinician agreement for cues. Table 9 presents the mean inter-clinician correlation (Phi and Porter Coefficients) for the four cases on pre- and posttest.

Table 9

Intercorrelations on Cues Requested

Simulated Case	Pretest				Posttest			
	Phi Coef.	Stan. Dev.	Porter Coef.	Stan. Dev.	Phi Coef.	Stan. Dev.	Porter Coef.	Stan. Dev.
1	.09	.19	.23	.12	.36	.14	.36	.09
2	.08	.17	.22	.10	.40	.12	.39	.10
3	.17	.12	.22	.07	.27	.11	.26	.08
4	.20	.19	.28	.14	.29	.15	.32	.10
Grand Mean	.13	.06	.24	.03	.33	.06	.33	.06

As the data in Table 9 indicate, students in this study seem to have had a higher agreement on cues they collected on the posttest, than on the pretest. This may suggest that training under the Model of Reading and Learning to Read structured the way in which the students gathered information about a case.

Analysis of Background Variables

In addition to finding the students' diagnostic agreement with each other, we also calculated correlations to find the students' reliability with themselves on pre- and posttests. Data indicated that great variability existed among students. The scores ranged from .16 to .53

for Case 1, -.03 to .35 for Case 2, -.03 to .38 for Case 3, and .24 to .60 for Case 4.

A question of interest to us was, what could be the factors which affect this variability? Are there any relationships between students' education and experience and their diagnostic agreement scores?

Analysis of a few such background variables, such as number of years of teaching experience, degree status, and number of reading courses taken, indicated no relationships whatsoever between these variables and students' diagnostic agreement scores as measured by the intra-clinician correlation. Table 10 summarizes this background information for Case 1. Appendix K is the demographic form filled out by the students at the beginning of the course, from which the background information was obtained.

Table 10

Relationship Between Educational/Experience Factors and Subject's Agreement with Him/Herself

		Simulated Case No. 1		
Subject Number	Phi Coefficient	Number of Years of Teaching Experience	Degree Status	Number of Reading Courses Taken
16	.17	7	M.A.	7
1	.21	20	Education Specialist	-
8	.29	2	Graduate Student	4
6	.33	3	M.A.	3
14	.35	6	M.A.	2
4	.50	9	Ph.D. Candidate	9
27	.53	2	B.A.	

Conclusions

The data described in this paper suggest that the students in this study improved their diagnostic agreement during the five-week reading diagnosis course.

In general, students had higher agreement scores on posttests than on pretests on both the Data Base and Causality Checklists, as evidenced by the proportional agreement score and the inter-clinician correlations. This suggests that by the end of the course diagnostic judgments about a case were more consistent across the class than they were at the beginning of the course.

Analysis of cues collected revealed that higher agreement scores were also obtained for the cues the students collected on posttests than on those collected on pretests, as well as a shift in emphasis toward decisions based on raw data rather than interpreted data.

These findings suggest that, throughout the course, the students learned to approach diagnosis in a systematic and model-oriented manner.

Since years of teaching experience, educational level, and background in reading related coursework appear to be independent of the students' diagnostic consistency, the improvement seen in this study can only be attributed to the training received in this course.

A problem posed by this study is that, although improvement of students' diagnostic agreement has been demonstrated, it is quite difficult to isolate those factors that contributed to this change. Is it because the students were trained from a Model of Reading and Learning to Read, because they practiced with simulated cases, or both? If both, which of the two had the greater effect?

Time is another variable to be considered. In five weeks the diagnostic agreement score on the Causality Checklist doubled (going from .16 to .38). Would students improve more in a longer (and/or more intensive) course?

The use of decision aids in writing diagnoses seems to be a definite asset in improving diagnostic consistency. This is evidenced by the fact that, in this study, pretest scores tended to be higher than those obtained by clinicians in other studies, in which decision aids were not used (see Table 1).

Recommendations

There is an obvious need for more training studies, highly controlled and executed, to better understand the various factors that contribute to improvement in diagnostic agreement, and the degree to which each of of them results in such improvement.

Studies should be designed to indicate improvement in validity as well as reliability. It is not sufficient to improve students' diagnostic agreement; their diagnoses should also be accurate. Although one cannot study validity before one attains reliability, it may be that both can be incorporated into the design of one study. Two such studies might be (1) to check students' diagnoses with those of an expert, and (2) to follow the students in their own classrooms or clinics.

Careful attention should be given to the checklists used in such studies. The vocabulary should be standardized on all checklists, and students should be trained to use the checklists, so that they transfer information from their written diagnoses to the checklists as accurately as possible.

Clinicians should use decision-aids (e.g., the Diagnostic/Prescriptive

Summary Form) and/or check with one another before reaching final diagnoses (and suggesting remediation). Group diagnosis may be more accurate than the diagnosis of any one clinician.

Reference Notes

1. Vinsonhaler, J.F. The consistency of reading diagnosis (Research Series No. 28). East Lansing, Michigan: Institute for Research on Teaching, Michigan State University, 1979.

2. Gil, D., Hoffmeyer, E., VanRoekel, J., & Weinshank, A. Clinical problem solving in reading: Theory and research (Research Series No. 45). East Lansing, Michigan: Institute for Research on Teaching, Michigan State University, 1979.

3. Norman, G., Barrows, H., Neufeld, V., & Feightner, J. A data base for analysis of the process of clinical problem solving. Ontario, Canada: McMasters University, 1976.

4. Sherman, G.B., Weinshank, A.B., & Brown, S. Training reading specialists in diagnosis (Research Series No. 31). East Lansing, Michigan: Institute for Research on Teaching, Michigan State University, 1979.

5. Sherman, G.B. A Model of Reading and Learning to Read. Michigan State University, East Lansing, Michigan. In preparation.

References

Elstein, A., Shulman, L., & Sprafka, S. Medical problem solving: An analysis of clinical reasoning. Harvard University Press, 1978.

Garland, L.H. Studies on the accuracy of diagnostic procedures. American Journal of Roentgenology, 1959, 82, 25-38.

Paton, B.C. The accuracy of diagnosis of myocardial infarction. A clinocopathologic study. American Journal of Medicine, 1957, 23, 761-768.

Vinsonhaler, J., Wagner, C., & Elstein, A. The nquiry heory, an information processing approach to clinical problem solving. Medinfo 77 Proceedings, North Holland Publishing Co., 1977. (Also available as IRT Research Series No. 1. East Lansing, Michigan: Institute for Research on Teaching, Michigan State University, 1978.)

Yerushalmy, J. Reliability of chest radiography in the diagnosis of pulmonary lesions. American Journal of Surgery, 1955, 89, 231-240.

Course Outline

Office Hours: Room 353, Erickson Hall

Phone: 353-8750

12:40-3:30 T, Th

ED 830E - Diagnosis of Reading Difficulties, Summer Term, 1979
G. Sherman - Instructor
Assisted by J. Van Roekel and J. Snoddy

Text: Ekwall, Diagnosis and Remediation of the Disabled Reader, Allyn and Bacon

Tests:
1.0 Durrell Analysis of Reading Difficulties: Examiners Kit Harcourt, Brace and World, Inc.
2.0 Dolch, (Sets 1 and 2) Popper Words (Optional)
3.0 Slosson Intelligence Test Kit, Slosson Educational Publication, East Aurora, New York (optional)

General Course Objective:

The overall objective of this course is to help you develop both the confidence and commitment to become a professional decision-maker who can accept that role as necessary for successful teaching of the atypical child; the child who is having trouble learning to read.

Specific Course Objectives

1. To describe reading problems in terms of four vital reading processes and their effecting factors as defined by the Model of Reading and Learning (MORAL);
2. To perform reading diagnosis using hypothesis-directed inquiry, that is, to select appropriate tests and/or observational techniques to confirm or disconfirm your hunches about the child's problems.
3. To accurately administer and score diagnostic and achievement tests which measure cognitive, affective, and perceptual behaviors;
4. To write a clear and concise diagnostic report to include the data base, diagnostic summary and remedial suggestions.

Course Requirements and Evaluation

1. Completion of pre and post diagnostic/prescriptive test.
2. Performance on a diagnostic/prescriptive report of a real a child selected by you. It is to include 1) all test protocols and all observations and/or tapes that were developed during your sessions with the child, 2) the final diagnostic summary and 3) the remedial suggestions. This report is due at the start of class on Tuesday, July 24, 1979.

3. Completion of four diagnoses and accompanying remediations using simulated cases of reading difficulty.
4. Completion dates of the above three requirements are listed in the course schedule.
5. Attendance at all class sessions. Failure to attend all class sessions will be cause for grade penalty.

MORAL Effecting Factors

	Vital Sign	Task Effectors	Learning Effectors	Child Effectors
Instant Word Recognition (IWR)	IWR Behavior at Grade Placement	-Visual Discrimination -Visual Memory -Association Strategies -Visual Sequencing	-Levels of Independent Contextual Practice -Levels of Corrective and Reinforcing Feedback -Level of Meaningfulness in Practice Materials	-Visual Acuity -Importance of Reading to Family, Peers and student -Language Experience -Self Concept -Instructional History -Level of Persistance
Decoded Word Recognition (DWR)	DWR Behavior at Grade Level	-S/S/A:Single Consonants -S/S/A: Consonant Clusters -S/S/A: Vowel Phonograms -Visual Chunking -Blending of Recognized Syllables -Adjustment of Blended Words -Integration of Isolated Decoding Skills for word identification	-Levels of Independent Contextual Practice -Levels of Corrective and Reinforcing Feedback -Level of Meaningfulness in Practice Material -Dull on Isolated Decoding Tasks	-Auditory Acuity -Importance of Reading to Family, Peers, and student -Language Experience -Self Concept -Instructional History -Level of Persistance
Fluent Message Segmentation (FMS)	FMS Behavior at Grade Placement FMS Behavior at IWR Level	-Use of Punctuation Signals -Use of Syntactic Signals -Use of Semantic Signals -Application of IWR Behavior -Application of DWR Behavior	-Levels of Independent Contextual Practice -Levels of Corrective and Reinforcing Feedback -Level of Meaningfulness in Practice Materials -Quantity of Rehearsal or Recording -Dull on Punctuation	-Importance of Reading to Family, Peers, Student -Language Experience -Self Concept -Instructional History -Level of Persistance
Retentive Comprehension	R.C. Behavior at Grade Placement R.C. Behavior at IWR Level	-Application of Meaningful Vocabulary -Application of Inference Skills	-Level of Independent Contextual Practice -Levels of Corrective and Reinforcing Feedback -Intent or Set for Interaction with Message	-Importance of Reading to Family, Peers and Student -Level of Cognitive Function -Language Experience -Self Concept -Instructional History -Level of Persistance

Developed by Dr. George B. Sherman © 1979

Student_Instructions_for_Pretest/Posttest

The purpose of the Diagnostic Performance Test is both to instruct you and to evaluate your progress as a diagnostician of reading disabilities. The Diagnostic Performance Test requires you to diagnose a simulated case (SIMCASE) of reading disability. Our SIMCASES are "simulated" only in the sense that the real child is not present. The data in the SIMCASE was obtained from real children with real reading disabilities.

TEST_MATERIALS The materials you need for this diagnostic performance test are as follows:

(1) The SIMCASE including:

- The List of Available Information
- The Referral Information
- The Box of Materials

(2) The Diagnosis Summary Form
(3) An Audio Tape Cassette Player

If you do not have any of these items stop the tape player and signal your instructor.

TEST_PROCEDURES
Let us review the procedures for the diagnostic performance; you will do the following:

(1) Read the initial contact(referral) information. Look at the sketch of the child and listen to the taped interview;
(2) Diagnose the SIMCASE by collecting whatever further data you wish;
(3) Write up your diagnosis and remediation using the MORAL Diagnostic and Prescriptive Summary Form

This has been a general description of the Diagnostic Performance Test. If you have any questions, stop the recorder and signal your instructor.

HOW_TO_USE_THE_DIAGNOSIS_FORM
Let us next consider how to use the MORAL Diagnostic/Prescriptive Summary Form. Pick up the form and look it over for a few minutes.

As you can see, the form provides a guide for writing your diagnostic/prescriptive judgments. You are asked to write one reading problem per page together with its causal factor(s) and remedial suggestions.

If you are ready to begin the diagnostic performance test, simply

continue. Otherwise stop the recorder and signal your instructor.

HOW TO USE A SIMCASE

Let's now consider the specifics of how to use a SIMCASE. The case consists of the following items:

(1) The Initial Contact (Referral) Information
(2) The List of Available Information
(3) The Box of Materials

The Initial Contact Information provides some basic introductory information about the child. It includes a referral statement, a sketch of the child, and a taped interview.

The Information Inventory is the list of materials collected about the case. Each item of information is described by a KEYWORD that refers to its title and location in the box (e.g.,keyword for Durrell Oral Reading test booklet is DUR3).

Now open the SIMCASE Box.
Note the KEYWORDS on the folders. These keywords correspond to those in the Information Inventory. To find information in the SIMCASE--locate the proper KEYWORD in the Inventory. Use this KEYWORD to find the proper folder in the box.

Stop the recorder and ask your instructor to check you out.

Now that you have practiced, you can begin to diagnose and remediate your SIMCASE.

Begin by opening the Initial Contact folder. Examine the picture and the information, and listen to the tape.

When you have finished with the Initial Contact material, signal your instructor. After he or she responds you may begin collecting information to diagnose your case. You will have 45 minutes in which to collect information. The proctor will tell you when your time is up. You will have an additional 45 minutes in which to fill out the MORAL Diagnostic/Prescriptive Summary Form and the Diagnostic Checklist(s).

Please do not return any of the materials to the box. Leave them on the table.
If you have questions or problems at any time, signal your instructor.

Relax. Do your best.

Appendix D

ASZ2.1: Pretest/Posttest
Student Record Form

1. Student Name_____

2. Student Experimental Number_____

3. Date/Time (A.M. P.M.)_____

4. SIMCASE used_____

5. If any problems occurred, please explain.

_____

_____

_____

_____

6. Cues collected by student (Use Inventory Keyword)

1._____	13._____	25._____	37._____	49._____
2._____	14._____	26._____	38._____	50._____
3._____	15._____	27._____	39._____	51._____
4._____	16._____	28._____	40._____	52._____
5._____	17._____	29._____	41._____	53._____
6._____	18._____	30._____	42._____	54._____
7._____	19._____	31._____	43._____	55._____
8._____	20._____	32._____	44._____	56._____
9._____	21._____	33._____	45._____	57._____
10._____	22._____	34._____	46._____	58._____
11._____	23._____	35._____	47._____	59._____
12._____	24._____	36._____	48._____	60._____

Appendix E

MORAL Diagnostic-Prescriptive Summary Form #1

Select important features which characterize this child's reading problem and analyze them using this guide.

1. Statement of INSTANT WORD RECOGNITION (SIGHT WORD) problem

A. Statement of a factor which is causing this problem

Procedures for remediating this factor

B. Statement of another factor which is causing this problem

Procedures for remediating this factor

C. Statement of another factor which is causing this problem

Procedures for remediating this factor

D. Statement of another factor which is causing this problem

Procedures for remediating this factor

2. Statement of DECODED WORD RECOGNITION problem

A. Statement of a factor which is causing this problem

Procedures for remediating this factor

B. Statement of another factor which is causing this problem

Procedures for remediating this factor

C. Statement of another factor which is causing this problem

Procedures for remediating this factor

D. Statement of another factor which is causing this problem

Procedures for remediating this factor.

3. Statement of FLUENT MESSAGE SEGMENTATION(language, phrasing, fluency) problem

A. Statement of a factor which is causing this problem

Procedures for remediating this factor

B. Statement of another factor which is causing this problem

Procedures for remediating this factor

C. Statement of another factor which is causing this problem

Procedures for remediating this factor

D. Statement of another factor which is causing this problem

Procedures for remediating this factor.

4. Statement of RETENTIVE COMPREHENSION problem

A. Statement of a factor which is causing this problem

Procedures for remediating this factor

B. Statement of another factor which is causing this problem

Procedures for remediating this factor

C. Statement of another factor which is causing this problem

Procedures for remediating this factor

D. Statement of another factor which is causing this problem

Procedures for remediating this factor.

MORAL_Data_Base_Checklist

	DATA BASE	Ade-quate	Inad-equate	Not obser	Irrel-evant
1	INSTANT WORD RECOGNITION				
2	DECODED WORD RECOGNITION LEVEL 1.0-3.9				
3	DECODED WORD RECOGNITION LEVEL 4.0-8.9				
4	FLUENT TEXT SEGMENTATION AT INSTANT WORD RECOGNITION LEVEL				
5	MESSAGE COMPREHENSION				
6	VISUAL DISCRIMINATION OF WORD FORMS				
7	VISUAL MEMORY OF WORD FORMS				
8	ASSOCIATION STRATEGIES APPLIED DURING INSTANT WORD RECOGNITION				
9	VISUAL SEQUENCING OF WORD FORMS				
10	SOUND SYMBOL ASSOCIATION FOR SINGLE CONSONANTS				
11	SOUND SYMBOL ASSOCIATION FOR HIGH FREQUENCY CONSONANT CLUSTERS				
12	SOUND SYMBOL ASSOCIATION FOR HIGH FREQUENCY VOWEL PHONOGRAMS				
13	SOUND SYMBOL ASSOCIATION FOR RARE VOWEL PHONOGRAMS				
14	SOUND SYMBOL ASSOCIATION FOR RARE CONSONANT CLUSTERS				
15	VISUAL CHUNKING FOR DECODED WORD RECOGNITION				
16	BLENDING OF RECOGNIZED SYLLABLE UNITS FOR DECODED WORD RECOGNITION				
17	ADJUSTMENT OF BLENDED WORD TO CONVENTIONALLY SPOKEN FORM				
18	USE OF PUNCTUATION SIGNALS IN WRITTEN				

	MATERIAL				
19	USE OF SYNTACTIC CUES				
20	USE OF SEMANTIC CUES				
21	APPLICATION OF MEANING VOCABULARY FOR THE REQUIREMENTS OF THE MATERIAL READ				
22	APPLICATION OF COGNITIVE PROCESSES INVOLVING INFERENCE BUILDING				
23	LEVELS OF INDEPENDENT CONTEXTUAL PRACTICE FOR NORMAL DEVELOPMENT OF INSTANT WORD RECOGNITION ACHIEVEMENT				
24	LEVELS OF INDEPENDENT CONTEXTUAL PRACTICE FOR NORMAL DEVELOPMENT OF DECODED WORD RECOGNITION ACHIEVEMENT				
25	LEVELS OF INDEPENDENT CONTEXTUAL PRACTICE FOR NORMAL DEVELOPMENT OF FLUENT TEXT SEGMENTATION ACHIEVEMENT				
26	LEVELS OF CORRECTIVE AND REINFORCING FEEDBACK FOR NORMAL DEVELOPMENT OF INSTANT WORD RECOGNITION ACHIEVEMENT				
27	LEVELS OF CORRECTIVE AND REINFORCING FEEDBACK FOR NORMAL DEVELOPMENT OF DECODED WORD RECOGNITION ACHIEVEMENT				
28	LEVELS OF CORRECTIVE AND REINFORCING FEEDBACK FOR NORMAL DEVELOPMENT OF FLUENT TEXT SEGMENTATION ACHIEVEMENT				
29	MEANINGFULNESS LEVEL FOR INSTANT WORD RECOGNITION PRACTICE				
30	MEANINGFULNESS LEVEL FOR DECODED WORD RECOGNITION PRACTICE				
31	MEANINGFULNESS LEVEL FOR FLUENT TEXT SEGMENTATION PRACTICE				
32	DRILL ON GRAPHEMIC SEGMENTATION AND LETTER SOUND ASSOCIATION FOR CODE GENERALIZATION				
33	QUANTITY OF REREADING TO INTEGRATE LANGUAGE BEHAVIOR AND WORD BEHAVIOR				

34	DRILL ON PUNCTUATION FOR INTEGRATION OF LANGUAGE PHRASING WITH WORD RECOGNITION			
35	INTENT OR SET FOR INTERACTION WITH THE MESSAGE BEING READ			
36	APPLICATION OF DECODING SKILLS FOR SINGLE SYLLABLE WORD IDENTIFICATION			
37	APPLICATION OF DECODING SKILLS FOR MULTI SYLLABLE WORD IDENTIFICATION			
38	APPLICATION OF INSTANT WORD RECOGNITION BEHAVIOR IN CONTEXT			
39	APPLICATION OF DECODED WORD RECOGNITION BEHAVIOR IN CONTEXT			
40	VISUAL ACUITY FOR WORD LETTER DISCRIMINATION AND EXTENDED READING PRACTICE			
41	AUDITORY ACUITY FOR LETTER AND PHONEMIC FEATURE DISCRIMINATION			
42	IMPORTANCE OF READING TO FAMILY GROUP			
43	IMPORTANCE OF READING TO PEER GROUP			
44	IMPORTANCE OF READING TO STUDENT			
45	LEVEL OF COGNITIVE FUNCTION FOR RETENTIVE COMPREHENSION			
46	INSTRUCTIONAL HISTORY OF CONDITIONS OF LEARNING			
47	SELF CONCEPT AS A READER LEARNER			
48	LEVEL OF PERSISTENCE UNDER FRUSTRATION OR DISINTEREST WITH RESPECT TO READING			
49	LANGUAGE EXPERIENCE			

MORAL_Causality_Checklist

Directions

Circle the statement numbers of those statements that characterize the current cases.

1. Inadequate instant word recognition is partially caused by insufficient independent reading practice

2. Inadequate instant word recognition is partially caused by. inadequate visual acuity.

3. Inadequate instant word recognition is partially caused by inadequate visual discrimination.

4. Inadequate instant word recognition is partially caused by inadequate visual memory.

5. Inadequate instant word recognition is partially caused by inadequate persistance in a frustrating situation.

6. Inadequate decoded word recognition is partially caused by inadequate Sound Symbol Association for vowel phonograms.

7. Inadequate decoded word recognition is partially caused by inability to use Sound Symbol Association for vowel phonograms in polysyllabic words.

8. Inadequate decoded word recognition is partially caused by lack of independent application of known decoding skills.

9. Inadequate decoded word recognition is partially caused by inadequate adjustment of blended words to conventionally spoken form.

10. Inadequate decoded word recognition is partially caused by inadequate blending of word parts.

11. Inadequate decoded word recognition is partially caused by lack of practic.

12. Inadequate fluent message segmentation(at Instant Word Recognition level) is partially caused by inappropriate use of punctuation.

13. Inadequate fluent message segmentation(at Instant Word Recognition level) is partially caused by deficit in language development.

14. Inadequate fluent message segmentation(at Grade Placement) is partially caused by inadequate Instant Word Recognition.

15. Inadequate fluent message segmentation(at Grade Placement) is partially caused by slow Decoded Word Recognition.

16. Inadequate fluent message segmentation(at Grade Placement) is partially caused by inadequate language development.

17. Inadequate fluent message segmentation(at Instant Word Recognition level) is partially caused by insufficient oral reading practice.

18. Inadequate fluent message segmentation(at Instant Word Recognition level) is partially caused by inadequate language development.

19. Inadequate retentive comprehension(at Instant Word Recognition level) is partially caused by inadequate inferential skills.

20. Inadequate retentive comprehension(at Grade Placement) is partially caused by inadequate inferential skills.

21. Inadequate retentive comprehension(at Grade Placement) is partially caused by inadequate language experience.

22. Insufficient independent reading practice is partially caused by lack of importance of reading to family.

23. Insufficient independent reader practice is partially caused by lack of meaningfulness level for practice.

24. Inadequate adjustment of blended words to conventionally spoken form is partially caused by lack of meaning vocabulary.

25. Auditory acuity deficit impairs overall phonics skills:both generalization and transfer performance.

Appendix H

DIAGNOSTIC CHECKLIST

Clinical Studies Project
Institute for Research on Teaching
Michigan State University

Clinician: _____

Session: _____

Case: _____

Date: _____

Time: _____

Location: _____

© Clinical Studies Project
IRT - 1979

	S	W	O		S	W	O
PHYSICAL				FAMILY			
Physical: General				Family: General			
Physical development				Home environment			
General health				Parent-school cooperation			
Vision: General				Parent educational background			
Vision: Acuity				Parent attitude toward school			
Hearing: General				Parent-child relationship			
Hearing: Acuity				Sibling relationship			
Speech articulation				Other			
Motor coordination							
Other							
				PERCEPTION			
				Perception: General			
				Auditory memory			
				Auditory discrimination			
				Auditory sequencing			
AFFECTIVE				Auditory blending			
Affective: General				Visual memory			
Attitude toward school				Visual discrimination			
Attitude toward reading: Instructional				Visual sequencing			
Attitude toward reading: Independent				Visual motor integration			
Relationships with peers				Other			
Motivation for reading							
Behavior in the classroom							
Emotional adjustment				LANGUAGE			
Self-concept				Language: General			
Variety of interests				Listening vocabulary			
Ability to deal with new situations				Speaking vocabulary			
Other				Verbal fluency			
				Other			

S - Strength; W - Weakness; O - Observation

	S	W	O		S	W	O
SCHOOL				**ORAL READING**			
School: General				Oral reading: General			
Grade level placement				Oral reading: Rate			
Ability to work in a tutoring situation				Oral reading: Fluency			
Ability to work in a small group				Oral reading: Punctuation			
Ability to apply reading skills				Oral reading: Self-correction			
Attending behavior				Oral reading: Phrasing			
Intellectual potential: General				Oral reading: Intonation			
Intellectual potential: Verbal				Substitutions contextually acceptable			
Intellectual potential: Nonverbal				Omissions contextually acceptable			
Ability to do grade level work				Insertions contextually Acceptable			
Ability to read at grade level				Use of context to determine word pronunciation			
Appropriateness of instructional materials				Independent reading level: Fluency			
Other				Instructional reading level: Fluency			
				Other			
STRUCTURAL ANALYSIS				**SILENT READING**			
Word Analysis: General				Silent Reading: General			
Structural Analysis: General				Silent Reading: Rate			
Use of root words				Silent Reading: Comprehension			
Use of suffixes				Other			
Use of prefixes							
Use of syllables							
Blending word parts into whole words							
Other							

S = Strength; W = Weakness; O = Observation

	S	W	O		S	W	O
PHONETIC ANALYSIS				WORD RECOGNITION			
Word analysis: General				Word recognition: General			
Phonetic analysis: General				Word recognition: Basic sight words			
Use of letter-sound association				Use of whole word approach			
Use of initial consonant sounds				Use of initial letters in word identification			
Use of final consonant sounds				Use of final letters in word identification			
Use of blends: General				Use of medial letters in word identification			
Use of consonant diagraphs				Use of letter order (reversals)			
Use of word families				Other			
Use of syllables							
Ability to blend sounds into whole word							
Other							
				COMPREHENSION			
				Comprehension: General			
				Comprehension: Oral			
				Comprehension: Silent			
				Comprehension: Listening			
OTHER				Comprehension: Vocabulary			
				Use of context to get meaning			
				Use of word order to get meaning			
				Recall of information			
				Comprehension: Use of Specific strategies			
				Other			

Appendix I

SIMCASE #1

Cues selected most frequently by 50 percent of subjects or more

Type of Information	Pre. Test Score	Pre. Examiner Comments	Pre. Test Booklet	Pre. Audio Record	Pre. Test Directions	Post Test Score	Post Examiner Comments	Post Test Booklet	Post Audio Record	Post Test Directions
Background Information										
Physical/Health		.71					1.00			
Home/Family		-					.86			
Classroom Information		-					.71			
Biographical Data		.57					-			
Diagnostic Test of Reading Difficulty (Durrell)										
Oral Reading	.71	.57	.57	-	-	-	-	.86	.57	-
Silent Reading	.71	.57	-		-	-	-	.86		-
Listening Comprehension	-	-	-		-	.57	-	.86		-
Word Recognition/Word Analysis	.71	.57	-	-	-	-	-	.86	-	-
Visual Memory of Words-primary	.71	.57	-		-	-	-	.86		-
Hearing Sounds in Words	.71	.57	-		-	-	-	.71		-
Basic Sight Vocabulary(Dolch)	.57	-	.57	-	-	-	-	.71	-	-
Graded Word List (SORT)	-	-	-	-	-	-	-	1.00	-	-
Cognitive Ability (WISC)										
Full Scale	.71	.57			-	-	-			-
Verbal Scale	.57	-			-	-	-			-
Performance Scale	.57	-			-	-	-			-
Diagnostic Test of Reading Difficulty (Gates-McKillop)										
Recognizing & Blending Common Word Parts	.57	-	.71	-	-	-	-	1.00	-	-
Achievement Test-Group (Gates-McGinitie)										
Vocabulary	-	-	-			-	-	.71		
Comprehension	-	-	-			-	-	.57		
Achievement Test-Individual (Peabody)										
Reading Recognition	-	.57	-		-	-	-	-		-
Reading Comprehension	.57	-	-		-	-	-	-		-

- indicates less than 50% agreement

SIMCASE #2

Cues selected most frequently by 50 percent of subjects or more

Types of Information	Pre: Test Scores	Pre: Examiner Comments	Pre: Test Booklet	Pre: Audio Record	Pre: Test Directions	Post: Test Scores	Post: Examiner Comments	Post: Test Booklet	Post: Audio Record	Post: Test Scores
Background Information										
Parent Form		1.00					.71			
Teacher Form		-					.57			
Diagnostic Test of Reading Difficulty (Durrell)										
Oral Reading	.57	.71	.57	-	-	-	-	1.00	.57	-
Silent Reading	-	.57	-		-	-	-	.86		-
Listening Comprehension	-	-	-		-	-	-	.71		-
Word Recognition/Word Analysis	-	-	-		-	-	-	1.00		-
Visual Memory of Words-Primary	-	.57	-		-	-	-	.71		-
Hearing Sounds in Words	.57	.57	.57		-	-	-	.86		-
Profile	.57					-				
Basic Sight Vocabulary (Dolch)	.57	.71	-		-	-	-	-		-
Graded Word List (SORT)	-	-	-		-		-	1.00		-
Cognitive Ability (WISC)										
Full Scale	.57	.57	-			-	-	-		
Performance Scale	.57	-	-		-	-	-	-		-
Achievement Test Group (Gates-McGinitie)										
Vocabulary	-	-	-			.71	-	.71		
Comprehension	-	-	-			.86	-	.71		
Individual Reading Analysis (IRA)										
Common Suffixes	-	-	-		-	-	-	.86		-
Consonant Blends	-	-	-		-	-	-	.71		.71
Syllabication	-	-	-		-	-	-	1.00		1.00
Discrimination Assessment (Wepman)										
Auditory	-		.57		-	.57		.57		-
Visual Acuity								.57		
Auditory Acuity			-					1.00		
Unfinished Sentences (UNF)			-					.86		
Writing From Dictation (WRI)		-	.57		-		-	-		-

- indicates less than 50% agreement

SIMCASE #3

Cues selected most frequently by 50 percent of subjects or more

Type of Information	Pre: Test Score	Pre: Examiner Comments	Pre: Test Booklet	Pre: Audio Record	Pre: Test Directions	Post: Test Scores	Post: Examiner Comments	Post: Test Booklet	Post: Audio Record	Post: Test Directions
Background Information										
Standard Achievement Test Scores	-		-				1.00	-		
Teacher Form		-					.71			
Diagnostic Test of Reading (Durrell)										
Oral Reading	.57	.86	-	-	-	-	-	.86	.86	-
Silent Reading	-	-	-		-	-	-	.57		-
Listening Comprehension	-	-	-		-	.57	-	-		-
Word Recognition/Analysis	-	.57	-		-	-	-	.86		-
Profile Chart	.71					-				
Graded Word List (SORT)	-	-	-		-	-	-	1.00		-
Diagnostic Test of Reading (GMK) Difficulty										
Recognizing & Blending Common Word Parts	.71	.71	-		-	-	-	.86		-
Nonsense Words	.57	.57	-		-	-	-	.57		-
Syllabication	.71	.57	-		-	-	-	-		-
Group Reading Achievement (GMG)										
Vocabulary	.57	-	-		-	.71	-	-		-
Comprehension	.57	-	-		-	.57	-	-		-
Speed & Accuracy	.57	-	-		-	.57	-	-		-
Auditory Acuity		.71	-				-	-		
Visual Acuity		.71	-				.57	-		
Cognitive Ability (SIT)	.71	-	-		-	-	-	-		-

- indicates less than 50% agreement

SIMCASE #4

Cues selected most frequently by 50 percent of subjects or more

Type of Information	Pre: Test Scores	Pre: Examiner Comments	Pre: Test Booklets	Pre: Audio Record	Pre: Test Directions	Post: Test Scores	Post: Examiner Comments	Post: Test Booklets	Post: Audio Record	Post: Test Directions
Background Information										
School Record			.71					.71		
Teacher Form			.86					.71		
School Information			-					.57		
Parent Form			.86					1.00		
Diagnostic Analysis of Reading Difficulty (Durrell)										
Oral Reading	.71	.71	-	-	-	.57	-	.86	-	-
Silent Reading	.57	.57	-	-	-	-	-	-	-	-
Listening Comprehension	.57	.57	-		-	-	-	-		-
Word Recognition/Analysis	-	.71	-	-	-	.57	-	.57	-	-
Hearing Sounds in Words-primary	.57	-	-		-	-	-	-		-
Visual Memory of Words-primary	-	.57	-		-	-	-	-		-
Intermediate Spelling List 1	.57	.57	-		-	-	-	-		-
Basic Sight Vocabulary (Dolch)	.86	.86	.71		-	.71	.71	.86		-
Graded Word List (SORT)	-	-	-		-	.86	-	.86		-
Cognitive Ability (WISC)										
Verbal Scale	.71				-	.86				-
Performance Scale	.57				-	.57				-
Full Scale	.57	-	-			.57	-	-		
Reading Diagnostic Test (Gates-McKillop)										
Recognizing & Blending Common Word Parts	.57	-	-	-	-	-	-	.71	-	-
Auditory Blending	-	.57	-	-	-	-	-	.57	-	-
Giving Letter Sounds	-	.57	-	-	-	-	-	-	-	-

- Indicates less than 50% agreement

SIMCASE #4 cont.

Cues selected most frequently by 50 percent of subjects or more

Type of Information	Pre				Post				
	Test Scores	Examiner Comments	Test Booklets	Audio Record	Test Directions	Test Scores	Examiner Comments	Test Booklets	Audio Record	Test Directions
Reading Achievement (Gates-McGinitie)										
Speed & Accuracy	-		-		-	.57		.57		-
Vision Test		.71	-				.71	.57		
Audiometric Recording		.57	-				.71	-		
Durrell Listening Reading Series (DLR) Intermediate Level										
Vocabulary	.57	-	-		-	-	-	-		-
Achievement Test (IOWA)										
Reading	-	-	-		-	.71	-	-		-

- indicates less than 50% agreement

Appendix J

Inter-Clinician Correlation

Phi

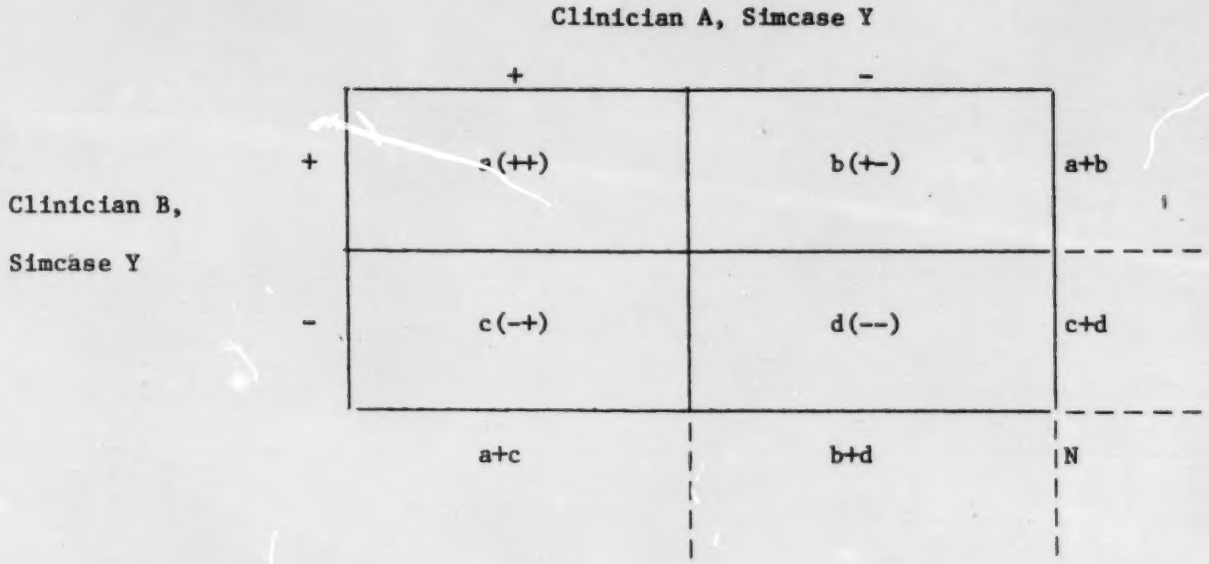
Given a domain for diagnoses/remediations/or cues (DX/RX or CX) for a given case, the Phi correlation is a measure of inter-clinician agreement. One Phi correlation is computed for each pair of clinicians.

The DX/RX/CX categories mentioned by one clinician are compared with those mentioned by a second clinician for the same case. This comparison is summarized in the table below.

CLINICIAN A, SIMCASE Y

CLINICIAN B, SIMCASE Y		PRESENT (+)	ABSENT (-)
	PRESENT (+)	Frequency count of statements in the domain present in both clinicians DX/RX or CX a	Frequency count of statements in the domain present in clinician B's session but not in clinician A's DX/RX/ or CX b
	ABSENT (-)	Frequency count of statements in the domain present in clinician A's session but not in B's DX/RX or CX c	Frequency count of statements in the domain absent in both clinicians' DX/RX or CX d

The calculation of the Phi correlation is derived from the table as follows:

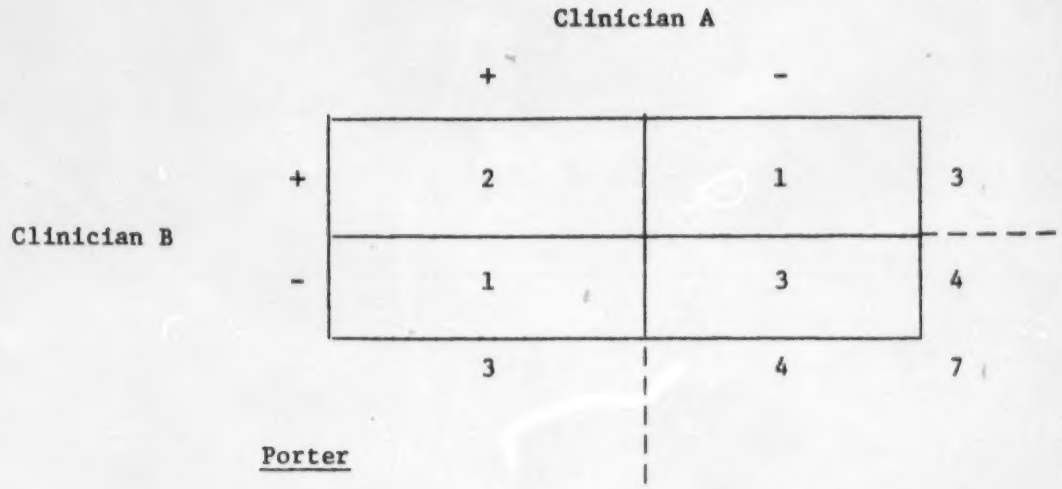


$$\text{Phi} = \frac{(axd - bxc)}{(a+c)x(b+d)x(c+d)x(a+b)}$$

The statistic is bounded by -1 (statements are in cells b and c only) and 1 (statements are in cells a and d only), only if the distribution in the marginals is equal. In all other cases the maximum and minimum values will be less than 1 and greater than -1.

An example of a completed table is as follows:

Statements of Clinician A, Simcase Y	Statements of Clinician B, Simcase Y	Domain of Statements
S1	S1	S1
S2	S2	S2
S3	S7	S3
		S4
		S5
		S6
		S7



Another statistic which is useful and based upon the 4-fold contingency table is the Porter Coefficient (P)* This coefficient is a measure of the agreement between two individuals' diagnoses or two diagnoses by the same individual as a function of the proportion of that agreement compared to the sum of the total agreements and non-agreements (i.e., the ++,+-,-+ boxes, and excluding the -- one.)

If A = no. categories common to both subjects or encounters,

B & C = no. categories mentioned by one subject (encounter) and not the other, then

$$P = \frac{A}{A + B + C}$$

This does not reflect agreement on not to select a category (diagnosis, cue, treatment) as is the case with the Phi correlation.

For the previous example, the Porter Coefficient would also be .67 as the total domain of categories is only 5. If the domain is large and the number of categories mentioned by both or either subject is small, the Phi and Porter values will differ, and the Porter coefficient will appear to be more reliable.

*developed by Dr. Andrew Porter, Institute for Research on Teaching, Michigan State University.

ED830E/SUMMER 1979 6/13/79

Student Information Form

Name:_____

Address:_____

_____

Phone No:_____

School:_____

Phone No:_____

Years taught (beyond student teaching):_____

Grade level(s) now teaching:_____

Subject(s) now teaching:_____

College:_____

Advisor:_____

Degree Status:_____

Major:_____

Previous Reading Courses (check if have taken or are now taking):

____325A, Mthd. Tch. Reading
____325B, Lang. Arts Elem. Grd.
____325C, Children's Literature
____483, Ind. Stud. Elem. Read.
____808, Mthd. Tch. Ch. Lit.
____809, Mthd. Tch. Lang. Arts
____830C, Elem. Read. Instr.
____830D, Read. Sec. Level
____830E, Diag. Read. Diff.
____830F, Prac. Remedial Read
____830H, Sem. Dev. Read.
____830J, Sem. Diag. Rem. Read

Other courses (check if have taken or are now taking):

____411, School Learning I
____464, Standardized Tests
____465, Testing and Grading
____811, School Learning II
____866B, WISC (WAIS) or Stanford Binet

Other Courses:

Course Name University

_____

_____

_____